Cell

# Tumor Archaeology Reveals that Mutations Love Company

Sunita R. Setlur[1] and Charles Lee[1,*]
[1]Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
*Correspondence: clee@rics.bwh.harvard.edu
DOI 10.1016/j.cell.2012.05.010

**Two studies by Nik-Zainal et al. and a study by Roberts et al. now provide new insights into the life span of human tumors and the mutational processes that shape the cancer genome. The bulk of a tumor's life span is shown to involve the development of subclones, many of which harbor mutations that are clustered in subchromosomal regions and appear to result from catastrophic mutational events.**

The classical theory of cancer genome evolution posits that cancer development occurs through gradual accumulation of mutations or traits that increase cell survival (Armitage and Doll, 1954; Stratton et al., 2009). However, some aspects of cancer evolution can occur on a much shorter timescale. For example, in the recently reported phenomenon, chromothripsis, individual chromosomes can be shattered and reassembled in a single catastrophic event (Stephens et al., 2011). In this issue of *Cell* and in a recent issue of *Molecular Cell,* Nik-Zainal et al. (2012a) and Roberts et al. (2012) describe another single catastrophic mutational event whereby point mutations accumulate rapidly at somatic hypermutation hotspots in a critical step of tumorigenesis.

In Nik-Zainal et al. (2012b), the mutational timeline of human cancers is first unveiled with the application of a newly developed, read-depth-based algorithm to identify patterns of mutations in each of 21 breast cancer genomes. First, the authors identified clusters of clones by measuring (1) read depth that was indicative of ploidy and (2) the allele fraction harboring the mutation (i.e., older mutations represent higher allele fractions). To formally simulate the clonal populations and identify the "most recent common ancestor," their approach clustered mutations according to the fraction of reads harboring the mutation and the frequency of the event in a given clone. This allowed for enumeration of distinct clonal populations inherent in the tumor. Interestingly, for genomes sequenced to ~40× mean coverage, the authors found that they had a 5% chance of identifying a clonal mutation present in 25% of the tumor sample. These findings illustrate that because of genetic heterogeneity, comprehensive identification of genetic variants in tumors by whole-genome DNA sequencing strategies requires much greater read depth than traditionally used for the identification of common germline/constitutional genetic variants (e.g., Mills et al., 2011).

Using an algorithm to phase mutations into haplotype structures, the authors found that driver mutations occur before the onset of large-scale chromosomal instability (Figure 1A). Shortly thereafter, and relatively early during tumorigenesis, a common ancestral clone is established from which subclones can subsequently be formed either by gradual accumulation of genetic alterations or as a result of catastrophic events including chromothripsis and "kataegis" (Greek for "shower" or "thunderstorm")—a phenomenon now revealed by Nik-Zainal et al. (2012a) and Roberts et al. (2012) whereby point mutations rapidly occur and cluster over hundreds ("microcluster") or millions ("macrocluster") of DNA bases. In Nik-Zainal et al. (2012a), five distinct kataegis "mutational signatures" were observed, each presumably generated by a biologically distinct mutational process (Figure 1B). Mutational signature A involved primarily C>T mutations at CpG sites while signature B was characterized by three kinds of mutations namely C>T transitions (at predominantly TpCpA and TpCpT sites), C>G mutations (at predominantly TpCpA and TpCpT sites) and C>A mutations (at TpCpA and TpCpT sites). Signature C is similar to signature A with respect to C>T substitutions at CpG sites but also had an elevated frequency of C>G mutations at CpG sites when compared with signature A. Signature D displayed a uniform distribution of the different mutational classes. Signature E had C>G mutations at TpCpA, TpCpC, and TpCpT trinucleotide sites but lacked the increased amount of C>T mutations at TpCp sites that are associated with signature B.

Roberts et al. (2012) also recognized this hypermutation phenomenon and experimentally investigated the underlying mechanism by inducing DNA damage of two artificially juxtaposed genes in the yeast *Saccharomyces cerevisiae*. The rate of mutations at the clustered regions was 500-fold higher than at other regions of the genome. The accumulation of mutations at TC(W)/(W)GA sites in a "strand-coordinated manner" (i.e., the mutations were clustered on the same parental DNA strand) was strong evidence for the occurrence of these mutational events in a single cell cycle. The authors further postulated that the clustering of these point mutations emanated from the formation of single-stranded DNA (ssDNA) that resulted from DNA double-strand break (DSB) repair or alternatively during formation of replication forks at the site of damaged DNA. Evidence for the latter mechanism came from experiments that disrupted genes crucial for the replication fork complex leading to replication fork defects as well as increased localized hypermutation.

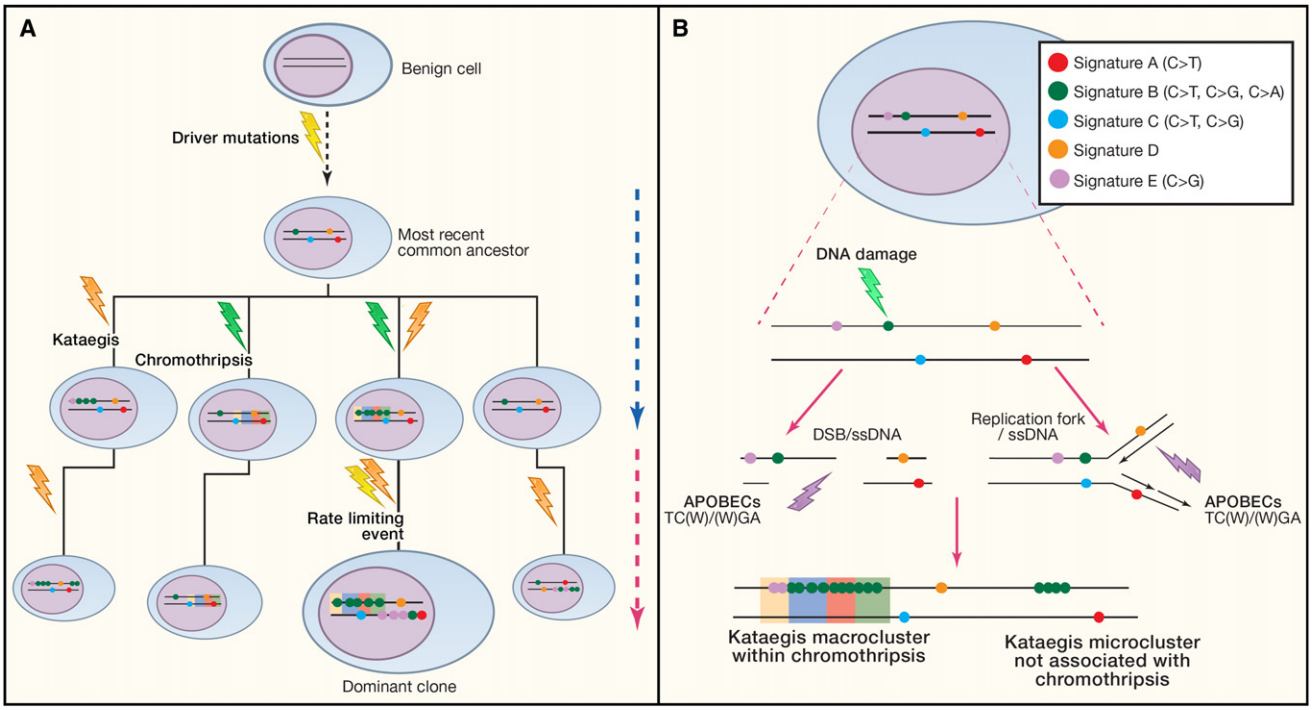


**Figure 1. Mutational Life Span of a Human Tumor**

(A) Whole-genome DNA sequencing provides archaeological evidence for the accumulation of point mutations and chromosomal rearrangements over decades of a tumor's life span. Driver mutations (such as *PIK3CA* and *TP53* mutations or *ERBB2, MYC*, or *CCND1* amplifications—yellow lightening) first appear and often subsequently lead to genomic instability. A "common ancestor" is then established relatively early during tumorigenesis. This is followed by a long period of tumor evolution (vertical blue line) where subclones are formed—in many cases, via chromosomal shattering and aberrant refusion events (chromothripsis—orange lightening) and/or regional accumulation of alkylation-based damage of cytosines and guanines (kataegis—green lightening). These events can occur concurrently or separately. A later rate-limiting step then occurs that permits one dominant clone to grow expeditiously (vertical red line) to account for more than 50% of a clinically recognizable tumor.

(B) Five distinct signatures (A–E) of kataegis were identified. Kataegis is thought to occur when DNA damages (light green lightening) result in double-strand DNA breaks that yield single-stranded DNA (ssDNA) or replication forks that expose ssDNA. The ssDNAs can be acted on by APOBEC enzymes with affinity for TC(W)/(W)GA motifs (purple lightening). Kataegis can deposit mutations within genomic clusters that are hundreds of bases in length (microcluster) or upwards of millions of bases in length (macrocluster) and can be associated with or independent of chromothripsis.

Based on the mutational patterns observed, both Roberts et al. (2012) and Nik-Zainal et al. (2012a) speculated that the point mutational clustering found in human tumors is mediated by the action of the APOBEC family of proteins. These proteins show a preference for ssDNA and mediate events such as the deamination of cytosines to uracil, which are then converted into thymine by base excision repair mechanisms.

Association of the chronological molecular timescale between chromothripsis and kataegis revealed that the two events could occur simultaneously at certain chromosomal regions but are not universal across the entire genome. In addition, kataegis was seen to occur multiple times during tumor evolution (Figure 1A). The findings from Nik-Zainal (2012b) also showed that a dominant subclone emerges from a late rate-limiting step, carrying multiple mutations that had accumulated since the "most recent common ancestor." The "most recent common ancestor" bore key driver mutations and appeared to be a relatively long-lived, quiescent population, which are also hallmarks of cancer stem cells. These features, coupled with an ability to generate new dominant clones, pose challenges for therapeutic intervention as resulting subclones are well poised to rapidly become resistant to new therapeutic interventions. Indeed, the presence of such competing subclones create a situation whereby each subclone is capable of gaining dominance quickly, even when an initial dominant subclone is eradicated from the tumor (Mullighan et al., 2008). In addition, generation of a new dominant clone may be facilitated by therapy-related cellular stress that could promote further chromothripsis/kataegis events leading to accelerated selection (Heng et al., 2010).

With more data accumulating from whole genome DNA sequencing of tumors, we now have an unprecedented opportunity to excavate data that can subsequently be used to reconstruct the mutational and evolutionary history of a cancer's genome. Although the majority of mutational events present in the cancer genome will likely not play a major role in tumor development, the unearthing of mutational signatures by using whole-genome DNA sequencing at high levels of sensitivity provides new insight into the processes that shape the cancer genome, offers opportunities to define new cancer subtypes, and ultimately contributes to a more comprehensive

understanding of disease progression. What were once thought to be merely random events in cancer evolution are now being recognized as important mutational signatures in the biography of a human tumor.

## REFERENCES


Armitage, P., and Doll, R. (1954). Br. J. Cancer *8*, 1–12.

Heng, H.H., Stevens, J.B., Bremer, S.W., Ye, K.J., Liu, G., and Ye, C.J. (2010). J. Cell. Biochem. *109*, 1072–1084.

Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K., et al; 1000 Genomes Project. (2011). Nature *470*, 59–65.

Mullighan, C.G., Phillips, L.A., Su, X., Ma, J., Miller, C.B., Shurtleff, S.A., and Downing, J.R. (2008). Science *322*, 1377–1380.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al. (2012a). Cell *149*, this issue, 979–993.

Nik-Zainal, S., Van Loom, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M., et al. (2012b). Cell *149*, this issue, 994–1007.

Roberts, S.A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., Klimczak, L.J., Kryukov, G.V., Malc, E., Mieczkowski, P.A., et al. (2012). Mol. Cell *46*, 424–435.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Cell *144*, 27–40.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). Nature *458*, 719–724.


# It Takes KASH to Hitch to the SUN


**Brian Burke**[1,*]

[1]Institute of Medical Biology, Singapore 138648, Singapore

*Correspondence: brian.burke@imb.a-star.edu.sg


DOI 10.1016/j.cell.2012.05.004

**LINC complexes are structures embedded within the nuclear envelope that mechanically couple the nucleus and cytoskeleton. They consist of SUN domain proteins of the inner nuclear membrane associated with KASH domain proteins in the outer nuclear membrane. Atomic resolution structures of SUN-KASH pairs now provide new insight in to the mechanisms of LINC complex assembly.**

Observations stretching back more than two decades have suggested that nuclei and nuclear components are mechanically coupled to the cytoskeleton. In multicellular organisms, this mechanical coupling may extend beyond the plasma membrane to the extracellular matrix and adjacent cells. More recently, studies on a variety of human diseases associated with defects in nuclear envelope (NE) proteins have revealed that, not only can the cytoskeleton affect nuclear organization, but changes in nuclear architecture may have a reciprocal affect on cytoskeletal function (Burke and Roux, 2009). In this issue, Sosa et al. (2012) provide an atomic resolution description of key interactions at the NE that link nuclear and cytoplasmic components.

The most prominent features of the NE are inner and outer nuclear membranes (INM and ONM) separated by a gap, or perinuclear space (PNS), of about 40–50 nm. The two membranes are spanned by nuclear pore complexes at annular junctions. The ONM also displays connections to the endoplasmic reticulum (ER). In this way, the ER, ONM, and INM represent separate domains within a single continuous membrane system, with the PNS forming an extension of the ER lumen. Whereas the INM contains a unique array of membrane proteins, the composition of the ONM closely resembles that of the ER. Nevertheless, the ONM is enriched in several integral membrane proteins that function as adapters for a variety of cytoskeletal components, including motor proteins. In vertebrate cells, these ONM proteins are represented, in part, by members of the Nesprin or Syne family.

The common feature of all known ONM-specific proteins, including Nesprins, is a C-terminal KASH (*K*larsicht, *A*nc1, *S*yne *h*omology) domain of 50–60 amino acid residues (Burke and Roux, 2009). The KASH domain consists of a single-membrane spanning helix followed by a sequence of about 30 residues that extends into the PNS. Localization of KASH domain proteins to the ONM is dependent upon interaction with integral proteins of the INM that belong to the SUN (Sad1p, Unc84) domain family. In mammalian cells, there are two widely expressed SUN domain proteins, Sun1 and Sun2. The SUN domain itself is a highly conserved ~200 amino acid C-terminal sequence that resides within the PNS at the end of a helical "stalk" and can interact directly with KASH domains of ONM proteins. In this way, proteins such as Sun1 and Sun2 function as transluminal tethers for Nesprin proteins in the ONM.

The nucleoplasmic domains of SUN proteins are associated with a variety of nuclear components, including the nuclear lamina, an important structural feature of the NE that is closely associated with both the INM and underlying